Chapter 17: Nonlinear regression

17.1: Polynomial Regression

- Sometimes the relationship between a dependent variable and a covariate has a non-linear relationship.

- If the curvature is not monotone (is increasing and decreasing), then inclusion of an $X^2$ term is probably a better option.

- Polynomial regression is a form of nonlinear regression.

- Possible ‘step-down’ procedure:
  - Include the highest order polynomial that appears needed after looking at the data
  - Fit the model (highest order and all lower orders)
  - Test the highest order term
  - If significant, stop. If not, remove highest term, re-fit the model, test next highest order. Continue until highest order is significant.
  - If $X^p$ is significant, you should include all lower order terms $X^{p-1}, X^{p-2}, \ldots, X^2, X$
Collinearity is often a problem:
- Situation improves if we use \((x_i - \bar{x})\) in place of \(x_i\). This is called ‘centering’.

Fitting a quadratic model with the original x-values:

\[
> \text{lm.out} = \text{lm}(y \sim x + I(x^2)) \\
> \text{vif(lm.out)}
\]

\[
x \quad I(x^2) \\
7.844986 \quad 7.844986
\]

Fitting a quadratic model with the centered x-values:

\[
> \text{lm.out.c} = \text{lm}(y \sim x.c + I(x.c^2)) \\
> \text{vif(lm.out.c)}
\]

\[
x.c \quad I(x.c^2) \\
1.019853 \quad 1.019853
\]

Much smaller variance inflation factors.

Centering does not fundamentally change the model, but may lead to computational benefits.

---

**Polynomial Regression with two predictors**

- The model:
  \[
  Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i}
  \]

- This model has the flexibility to fit a mean surface such as...

- Response surface methodology can be used to investigate the shape of a response surface.

  It considers interactions and quadratic effects.

  It can help find optimal process settings.

- Interactions of continuous variables can be difficult to interpret, though they can help model a surface.
**Example:** Life of a power cell

A researcher studied the effects of the charge rate and temperature on the life of a new type of power cell in a preliminary small-scale experiment.

**VARIABLES:**
- **Number of Cycles** -- life of the power cell, number of discharge-charge cycles before failure
- **Charge Rate** -- three levels (0.6, 1.0, 1.4 amperes)
- **Temperature** -- three levels (10, 20, 30 Celsius)

The researcher was not sure on the response function (or surface) in the region of the factors being studied, so she fit a second-order polynomial regression model.


---

The overall F-test is rejected, but none of the individual terms are significant.

It turns out, the VIFs from the model are quite large.

Model fit with centered predictor data:

Model fit with centered predictor data:

The p-value is the same in the centered model, but there is much less multicollinearity, and two predictors are significant.

A partial F-test suggests the group of second-order terms is not significant...

so, the reduced model surface was the accepted surface.
The centered variable fitted model:
\[ Y_i = 172 - 139.583 \cdot x_{1i} + 7.55 \cdot x_{2i} \]

Converting back to pre-shifted values:
\[ Y_i = 172 - 139.583 \cdot (X_{1i} - \bar{X}_1) + 7.55 \cdot (X_{2i} - \bar{X}_2) \]
\[ = 172 - 139.583 \cdot (X_{1i} - 1) + 7.55 \cdot (X_{2i} - 20) \]
\[ = 172 + 139.583 - 151 - 139.583 \cdot X_{1i} + 7.55 \cdot X_{2i} \]
\[ = 160.583 - 139.583 \cdot X_{1i} + 7.55 \cdot X_{2i} \]